Part 3: Critical Thinking (20 points)

Ethics & Bias (10 points):

Q: How might biased training data affect patient outcomes in the case study?

A: Biased training data can lead to unequal or inaccurate treatment recommendations across different patient groups. For instance, if the data underrepresents certain demographics (like women, elderly, or minority ethnic groups), the model may produce less accurate diagnoses or risk assessments for those groups. This could result in delayed treatment, misdiagnoses, or even life-threatening outcomes, ultimately reinforcing existing healthcare disparities.

Q: Suggest 1 strategy to mitigate this bias.

A: One effective strategy is to perform data auditing and rebalancing. This involves identifying underrepresented groups in the dataset and either collecting more data from these populations or applying techniques like resampling or reweighting to ensure that all groups are fairly represented during training. This helps improve fairness and ensures the model generalizes well across diverse patient populations.

Trade-offs (10 points):

Q: Discuss the trade-off between model interpretability and accuracy in healthcare.

A: In healthcare, model interpretability is crucial because doctors must understand and trust the model's decisions-especially in high-stakes environments like diagnosis or treatment planning. However, more interpretable models (like decision trees or logistic regression) may have lower accuracy compared to complex models (like deep neural networks). On the other hand, highly accurate models often act as 'black boxes,' making it difficult to justify clinical decisions. This trade-off means that in practice, healthcare systems must carefully balance accuracy with transparency to ensure both effectiveness and ethical accountability.

Q: If the hospital has limited computational resources, how might this impact model choice?

A: Limited computational resources would constrain the hospital to use simpler, less

resource-intensive models (e.g., logistic regression, random forest with fewer trees) rather than large-scale deep learning models. While these simpler models might not achieve the highest accuracy, they offer faster inference, require less memory, and are easier to deploy and maintain-making them more practical in low-resource settings.